Performance vs. Privacy: Evaluating the Performance of Predicting Second Primary Cancer in Lung Cancer Survivors with Privacy-preserving Approaches

Jui-Fu Hong
Department of Information Management
National Central University
Taoyuan, Taiwan
qqq951009@g.ncu.edu.tw

Yi-Ju Tseng
Department of Computer Science
National Yang Ming Chiao Tung University
Hsinchu, Taiwan
yjtseng@nycu.edu.tw
0000-0002-1814-5553

2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI) | 978-1-6654-8791-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/BHI56158.2022.9926935

***Abstract*—Deep learning has been widely used in the medical field to support medical decision making. Simultaneously, with the rise of data privacy protection, accessing clinical records across different institutions has become a possible challenge. Several approaches, such as federated and transfer learning, have been proposed to train models without accessing all the records from each institution, but the performance of these privacy-preserved models may not be as good as centralized approaches, which aggregate all records to build a centralized model. To explore the potential of privacy-preserving second primary cancer (SPC) prediction of lung cancer survivors using real-world data, we evaluated the performance of federated learning, transfer learning, learning with a single institution, and traditional centralized learning. We trained machine learning models using data from four hospitals and compared the model performances of learning from a single institution, centralized learning, federated learning, and transfer learning. The results show that federated learning outperformed other learning strategies in three of the four sites (AUROC from 0.733 to 0.777). However, only Site 6 showed that federated learning significantly outperformed all the other learning strategies (P < 0.05). In summary, federated learning can develop a unified model for the multiple institutions while maintaining data security.**

Keywords—machine learning, federated learning, transfer learning, second primary cancer, lung cancer

I. Introduction

Recently, machine learning and deep learning have been widely used in the medical field to support medical decision making [1], [2]. However, collecting large and diverse clinical records for model training is difficult. Most studies compensated for the data insufficiency by collecting and aggregating clinical records across multiple institutions, an approach known as centralized learning [3]. Concurrently, with the rise of data privacy issues and the establishment of strict regulations to protect personal information (e.g., Health Insurance Portability and Accountability Act in the US and General Data Protection Regulation in the European Union), accessing clinical records across institutions has become a potential issue. Due to these privacy concerns, the commonly used method is learning with a single institution dataset. Usually, the institution develops a prediction model to predict certain outcomes. A major disadvantage of learning with a single institution is that the smaller institution usually has less data and less diversity. Besides, it may hinder the model development process which can further cause negative impacts on the performance during the prediction. Several approaches, such as federated learning [4] and transfer learning [5], were introduced to solve these problems. These methods do not need to access all the clinical records from each institution, but only transfer model parameters, which can effectively prevent data leakage. These approaches are critical in certain fields, but the performance of these privacy-preserved models maybe not be as good as centralized approaches. Using real-world data, we should compare these methods with learning with a single institution and traditional centralized learning to explore the potential of privacy-preserving second primary cancer (SPC) prediction of lung cancer survivors.

II. Related Works

Federated learning was first proposed by Google [6]. To improve the query suggestion, they trained a decentralized model for Gboard. With the help of distributive characteristics, medical research increasingly leveraged federated learning for model training across multiple hospitals. Federated learning is often used mainly because of its new training paradigm, which addresses the problems of distributed and fragmented medical data. According to Jie Xu et al. [7], the models at each site are trained simultaneously. Afterward, it aggregated and returned the new model parameters to all the sites in each round. The whole process only transports the learned model parameters, thus avoiding the need for data sharing. Moreover, some studies compared the training differences between federated learning, centralized learning and learning from a single site. For example, Ittai Dayan et al. [8] compared models trained using these methods with electronic health record (EHR) and chest X-rays across 20 multinational hospitals, as well as the performance of these methods. In addition, the trained model was used to predict the oxygen requirement of each patient. The waste of medical resources can be effectively avoided by giving priority care and treatment to those patients who have a higher oxygen requirement. In the image segmentation task, Micah J. Sheller et al. [9] add Cyclic Institution Incremental learning and Institution Incremental learning into comparison. Although learning from a single site can probably achieve

Corresponding author: Y-J Tseng (yjtseng@nycu.edu.tw)

TABLE I. PATIENT CHARACTERISTIC

		site 2, n = 831	site 3, n = 7673	site 6, n = 1559	site 8, n = 3849
class, n (%)	No Second Primary Cancer	810 (97.5)	7447 (97.1)	1503 (96.4)	3757 (97.6)
	Second Primary Cancer	21 (2.5)	226 (2.9)	56 (3.6)	92 (2.4)
gender, n (%)	male	514 (61.9)	4330 (56.4)	953 (61.1)	2131 (55.4)
	female	317 (38.1)	3343 (43.6)	606 (38.9)	1718 (44.6)
age, n (%)	21 - 30	1 (0.1)	27 (0.4)	1 (0.1)	15 (0.4)
	31 - 40	8 (1.0)	241 (3.1)	25 (1.6)	114 (3.0)
	41 - 50	68 (8.2)	913 (11.9)	121 (7.8)	423 (11.0)
	51 - 60	168 (20.2)	1876 (24.4)	314 (20.1)	1010 (26.2)
	61 - 70	226 (27.2)	2164 (28.2)	410 (26.3)	1165 (30.3)
	71 - 80	257 (30.9)	1842 (24.0)	529 (33.9)	881 (22.9)
	81 - 90	96 (11.6)	581 (7.6)	150 (9.6)	231 (6.0)
	>=91	7 (0.8)	29 (0.4)	9 (0.6)	10 (0.3)
smoking behavior, n (%)	missing	308 (37.1)	3216 (41.9)	491 (31.5)	1547 (40.2)
	no smoking	179 (21.5)	2304 (30.0)	403 (25.8)	1196 (31.1)
	smoking	189 (22.7)	1577 (20.6)	418 (26.8)	745 (19.4)
	unknow	155 (18.7)	576 (7.5)	247 (15.8)	361 (9.4)
drinking behavior, n (%)	missing	308 (37.1)	3216 (41.9)	491 (31.5)	1547 (40.2)
	no drinking	273 (32.9)	3021 (39.4)	605 (38.8)	1493 (38.8)
	drinking	94 (11.3)	811 (10.6)	195 (12.5)	397 (10.3)
	unknow	156 (18.8)	625 (8.1)	268 (17.2)	412 (10.7)
betel nut chewing behavior, n (%)	missing	308 (37.1)	3216 (41.9)	491 (31.5)	1547 (40.2)
	no chewing	340 (40.9)	3520 (45.9)	663 (42.5)	1688 (43.9)
	chewing	27 (3.2)	335 (4.4)	133 (8.5)	183 (4.8)
	unknow	156 (18.8)	602 (7.8)	272 (17.4)	431 (11.2)
surgical margins of the primary site, n (%)	missing	1 (0.1)	47 (0.6)	229 (14.7)	1 (0.0)
	0	167 (20.1)	1551 (20.2)	9 (0.6)	600 (15.6)
	1 ~ 6	10 (1.2)	34 (0.4)	1321 (84.7)	15 (0.4)
	unknow	653 (78.6)	6041 (78.7)	0 (0.0)	3233 (84.0)
clinical stage, n (%)	0	0 (0.0)	29 (0.4)	7 (0.4)	16 (0.4)
	I	125 (15.0)	1194 (15.6)	210 (13.5)	455 (11.8)
	II	42 (5.1)	260 (3.4)	67 (4.3)	101 (2.6)
	III	209 (25.2)	1557 (20.3)	357 (22.9)	748 (19.4)
	IV	395 (47.5)	3994 (52.1)	789 (50.6)	2140 (55.6)
	unknow	60 (7.2)	639 (8.3)	129 (8.3)	389 (10.1)
pathologic differentiation, n (%)	missing	0 (0.0)	9 (0.1)	43 (2.8)	5 (0.1)
	well	62 (7.5)	1145 (14.9)	90 (5.8)	112 (2.9)
	moderately	122 (14.7)	1424 (18.6)	206 (13.2)	505 (13.1)
	poorly	149 (17.9)	1668 (21.7)	397 (25.5)	291 (7.6)
	undifferentiated	3 (0.4)	32 (0.4)	21 (1.3)	9 (0.2)
	unknow	495 (59.6)	3395 (44.2)	802 (51.4)	2927 (76.0)
positive regional lymph nodes, n (%)	0	125 (15.0)	1379 (18.0)	203 (13.0)	502 (13.0)
	1 ~ 2	23 (2.8)	174 (2.3)	37 (2.4)	89 (2.3)
	3 ~ 6	11 (1.3)	101 (1.3)	18 (1.2)	41 (1.1)
	7 ~ 15	6 (0.7)	44 (0.6)	10 (0.6)	25 (0.6)
	> =16	4 (0.5)	6 (0.1)	2 (0.1)	4 (0.1)
	unknow	662 (79.7)	5969 (77.8)	1289 (82.7)	3188 (82.8)
surgical resection, n (%)	no date	647 (77.9)	6028 (78.6)	1315 (84.3)	3208 (83.3)
	has date	184 (22.1)	1645 (21.4)	244 (15.7)	641 (16.7)
tumor size, n (%)	missing	1 (0.1)	1 (0.0)	15 (1.0)	3 (0.1)
	1 - 49mm	381 (45.8)	3944 (51.4)	818 (52.5)	1855 (48.2)
	50 - 99mm	169 (20.3)	1750 (22.8)	375 (24.1)	867 (22.5)
	100 - 149mm	21 (2.5)	156 (2.0)	26 (1.7)	64 (1.7)
	>= 150mm	0 (0.0)	11 (0.1)	2 (0.1)	2 (0.1)
	unknow	259 (31.2)	1811 (23.6)	323 (20.7)	1058 (27.5)
radiology treatment , n (%)	no date	652 (78.5)	5814 (75.8)	1135 (72.8)	3126 (81.2)
	have date	179 (21.5)	1859 (24.2)	424 (27.2)	723 (18.8)
systemic therapy, n (%)	no date	326 (39.2)	2583 (33.7)	602 (38.6)	1128 (29.3)
	have date	505 (60.8)	5090 (66.3)	957 (61.4)	2721 (70.7)
BMI, n (%)	missing	308 (37.1)	3220 (42.0)	491 (31.5)	1547 (40.2)
	<18.5	28 (3.4)	182 (2.4)	47 (3.0)	116 (3.0)
	18.5-23.9	177 (21.3)	1785 (23.3)	372 (23.9)	922 (24.0)
	>=24	159 (19.1)	1641 (21.4)	352 (22.6)	824 (21.4)
	unknow	159 (19.1)	845 (11.0)	297 (19.1)	440 (11.4)
separate tumor nodules in ipsilateral lung, n (%)	missing	463 (55.7)	3662 (47.7)	747 (47.9)	1822 (47.3)
	no	313 (37.7)	2765 (36.0)	554 (35.5)	1247 (32.4)
	yes	55 (6.6)	1108 (14.4)	258 (16.5)	715 (18.6)
	unknow	0 (0.0)	138 (1.8)	0 (0.0)	65 (1.7)
visceral pleural invasion / elastic layer, n (%)	missing	309 (37.2)	3081 (40.2)	491 (31.5)	1468 (38.1)
	no	36 (4.3)	912 (11.9)	100 (6.4)	240 (6.2)
	yes	49 (5.9)	304 (4.0)	85 (5.5)	184 (4.8)
	unknow	437 (52.6)	3376 (44.0)	883 (56.6)	1957 (50.8)
assessment of performance status before treatment, n (%)	missing	463 (55.7)	4201 (54.8)	779 (50.0)	1886 (49.0)
	0	120 (14.4)	1179 (15.4)	76 (4.9)	513 (13.3)
	1	157 (18.9)	1866 (24.3)	552 (35.4)	1089 (28.3)
	2	45 (5.4)	253 (3.3)	77 (4.9)	190 (4.9)
	3	12 (1.4)	104 (1.4)	22 (1.4)	54 (1.4)
	4	15 (1.8)	30 (0.4)	13 (0.8)	12 (0.3)
	unknow	19 (2.3)	40 (0.5)	40 (2.6)	105 (2.7)
malignant pleural effusion, n (%)	missing	309 (37.2)	3220 (42.0)	491 (31.5)	1547 (40.2)
	no	141 (17.0)	1361 (17.7)	330 (21.2)	812 (21.1)
	yes	93 (11.2)	795 (10.4)	148 (9.5)	467 (12.1)
	unknow	288 (34.7)	2297 (29.9)	590 (37.8)	1023 (26.6)
EGFR (Epidermal growth factor receptor) gene mutation, n (%)	missing	463 (55.7)	3807 (49.6)	747 (47.9)	1887 (49.0)
	no	85 (10.2)	735 (9.6)	150 (9.6)	439 (11.4)
	yes	96 (11.6)	1478 (19.3)	307 (19.7)	796 (20.7)
	unknow	187 (22.5)	1653 (21.5)	355 (22.8)	727 (18.9)

good results when predicting the internal testing set, it does not generalize well to the external dataset. However, the validation results of Ittai Dayan et al. [8] and Micah Sheller et al. [9] show that whether in internal or external validation datasets, federated learning outperforms only learning with a single institution in most medical institutions, and it also has better generalization. Although each institution's data were strictly regulated, federated learning can still train models by aggregating model parameters from each site and producing comparable results to centralized learning.

However, federated learning has limitations. For example, the difference in demographics, instrumentation used, or other factors may cause bias, also known as non-IID (independent and identically distributed) problems. For one thing, those federated learning participants with a non-IID problem may cause the performance of the final trained model to be a side effect. Furthermore, the federated learning model may not accurately predict the heterogeneous institutional dataset. Therefore, Tian Li et al. has proposed a special aggregation method called FedProx [10], which successfully mitigate non-IID problems.

Lung cancer is one of the most common cancers and has been the leading cause of cancer-related death in Taiwan, the United States, and worldwide [11]. As the incidence of lung cancer soars, the number of patients diagnosed with SPC after lung cancer is also rising. Like recurrence, early detection of SPC risk is critical to improving prognosis. Previous studies only used a single institution or centralized learning to developed machine learning models to predict SPC [12]–[14]. Therefore, we should develop and evaluate the SPC risk prediction models for lung cancer survivors using privacy-preserved machine learning technologies and compare the performance with the traditional approaches.

III. METHODS

A. Data Sources and Clinical Settings

The design is a retrospective cohort study on the largest multi-institutional database that includes de-identified electronic medical records from the Chang Gung Memorial Hospitals between 2010 and 2018. We collected data from the Keelung (Site 2), Linkou (Site 3), Chiayi (Site 6), and Kaohsiung (Site 8) CGMHs, which provide both primary and tertiary medical care in northern and southern Taiwan. The data used to develop the SPC prediction models are age, gender, and cancer registry information, including separate tumor nodules in the ipsilateral lung, visceral pleural invasion/elastic layer, assessment of performance status before treatment, malignant pleural effusion, EGFR (Epidermal growth factor receptor) gene mutation, pathologic differentiation, tumor size, positive regional lymph nodes, clinical stage, surgical margins of the primary site, body mass index (BMI), surgical resection, radiology treatment, systemic therapy, smoking behavior, betel nut chewing behavior, and drinking behavior. Patients diagnosed with lung cancer who had followed up for at least 6 months were considered lung cancer survivors. The Chang Gung Medical Foundation Institutional Review Board approved this study (IRB no. 201901386B0) and waived the requirement for patient consent.

B. Dataset and Preprocess

Our dataset included four hospitals, and each was considered an independent site. We drop the records with over 10 null values, and after that, we use the median to fill the remaining null values.

All the features were coded in categorical format. The sparse matrix might cause the curse of dimensionality by using one-hot encoding. Thus, we used target encoding as an alternative. Each site randomly split its dataset into two parts: 80% for training and 20% for testing.

C. *Model Training Strategy*

We compare four machine learning strategies, including learning with a single institution, centralized learning, transfer learning, and federated learning. When learning with a single institution, models were trained on each site using only its local training set. Centralized learning collects and aggregates all sites' local training set and trains one universal model. In transfer learning, the site with the largest local training set is used to build a pretrained model, and the other sites finetune the pretrained model using their own local training sets. In federated learning, each site receives an initial parameter from the central server, and then each site trains the model using its local training set for a number of epochs. After training the model, each site's model parameters are transferred to the central server. Then, the central server aggregates all the model parameters to generate a new model and sends it back to each site. This process is iterated over several rounds until it finishes. In our experiment, we used "Flower" [15] as the federated learning framework, which can simulate the client and server architecture of the federated learning system on a single machine without needing a multi-machine system. Furthermore, we set the federated rounds at 20, with 5 local training epochs per round for each site. Thus, the training epoch of the other three methods was set to a hundred to correspond to the total epochs of federated learning. In addition, models were evaluated on the testing set holding out from each site. All sites joined the training process in each round of federated learning. The weighted average (i.e., FedAvg [4]) was used as the aggregation strategy when the model parameters were sent to the central server for aggregation. All the training methods mentioned above were simulated 57 times—we got 57 evaluation results for each training method and further averaged 57 results to get the mean value. To compare the results of each training method, we conducted the Kruskal-Wallis test and Dunn's multiple comparison post hoc test on site 2 and site 6, one-way ANOVA test and Bonferroni multiple comparison post hoc test on site 3 and site 8.

D. *Model Architecture*

A Deep Neural Network (DNN) is used in our experiment. The model architecture includes an input layer, two hidden layers with 16 and 10 neurons, respectively, and an output layer with a sigmoid activation function. Due to the imbalanced class problem in the dataset of each site, we choose Focal Loss [16] as our model loss function, adding the modulating factor to the cross-entropy to focus learning on hard misclassified examples. Furthermore, we train the model using the Adam optimizer with a learning rate of 0.001. The models output a continuous value between 0 and 1, and the metric we used is the area under the receiver operating characteristic curve (AUROC).

IV. RESULT

Table 1 shows the patients' characteristics of each site, and Table 2 and Figure 1 show the evaluation results of different training strategies and sites. The result indicates that in site 2, federated learning models outperformed (0.764 ± 0.153) centralized models (0.543 ± 0.207, $P < 0.001$) but showed no significant difference between learning from a single site (0.761 ± 0.146). In site 6, federated learning (0.738 ± 0.066) statistically and significantly outperformed other training methods ($p < 0.05$).

Although the federated learning models did not significantly outperform the other models at site 8, it did slightly outperform other training methods. Finally, in site 3, which has the largest dataset, the centralized learning method performed the best (0.784 ± 0.034) but showed no significant difference between federated learning.

We found that the prediction results in site 2 were widely distributed for all learning strategies, and the centralized learning did not outperform the other three training methods. However, the predicting results of site 3, which has the biggest dataset, were more concentrated on each learning strategy.

TABLE I. PERFORMANCE OF PREDICTING SECOND PRIMARY CANCER IN LUNG CANCER SURVIVOR

Machine learning methods	Area under the receiver operating characteristic curve (AUROC) for each hospital, mean (SD)			
	Site 2	Site 3	Site 6	Site8
Learning from a Single Site	0.761 (0.146)	0.774 (0.034)	0.706 (0.067)	0.706 (0.070)
Centralized Learning	0.543 (0.207)	**0.784** (0.034)	0.708 (0.070)	0.715 (0.070)
Transfer Learning	0.735 (0.142)	0.774 (0.034)	0.693 (0.067)	0.712 (0.070)
Federated Learning	**0.764** (0.153)	0.777 (0.032)	**0.738** (0.066)	**0.733** (0.069)

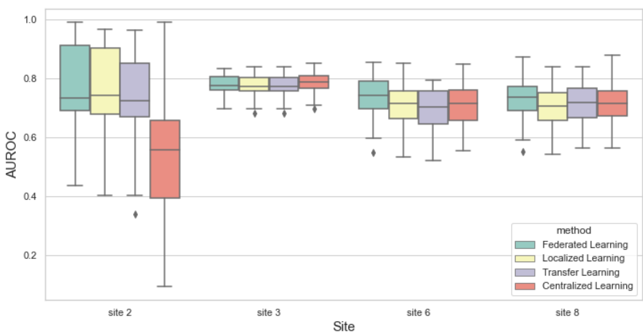

Fig. 1. PERFORMANCE OF PREDICTING SECOND PRIMARY CANCER IN LUNG CANCER SURVIVORS. AUROC*:* Area under the receiver operating characteristic curve

V. DISCUSSION

In 3 of the 4 sites, federated learning outperformed centralized learning. In particular, in each site, the performance of federated learning models was better compared to learning with a single institution. It proved that the model could perform similarly to the centralized learning method through the distributed, privacy-preserving training method without collecting data from other sites.

Out of the 4 datasets, sites 2 and 3 are the smallest and largest datasets, respectively. The data size may affect the distribution of the evaluation performance. For example, the evaluation results were widely distributed in site 2. However, the evaluation results of site 3 were more concentrated than the others. Hence, the smaller dataset may make the predicted results more unstable in each learning method.

There are some limitations to our study. First, we only included data sets from four hospitals in Taiwan, and these four institutions belong to the same hospital groups. Thus, the results might not be applied to the other datasets. Moreover, due to race or other demographic baseline differences, the final model built with federated learning may not work well on the external datasets. The dataset is then highly imbalanced. We have tried to solve the class imbalance problem by using synthesized minority oversampling technique (SMOTE) [17] oversampling while training a federated learning model. However, it deteriorates the final result when evaluated. Finally, the federated learning framework we used is now not compatible with traditional machine learning algorithms. Therefore, we only use the basic DNN model for the research. It may not be the most suitable model for these datasets. In the future, we will apply different machine learning methods with a federated learning framework.

VI. Conclusion

Through private preserving machine learning strategies, the patient data can be completely protected and mitigate the risk of data leakage without compromising the prediction performance of SPC for lung cancer survivors. Although there are some limitations to federated learning, with continuous improvement, this training paradigm can be applied to different diseases and develop a general and robust model for the clinical institution.

Acknowledgement

This study was supported by grants from the Ministry of Science and Technology, Taiwan (MOST 111-2636-E-A49-014 and 111-2628-E-A49-026-MY3). These institutions had no role in the study design, decision to publish, or preparation of the manuscript.

References

[1] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine," *N. Engl. J. Med.*, vol. 380, no. 14, pp. 1347–1358, Apr. 2019, doi: 10.1056/NEJMra1814259.

[2] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," *Nat. Med.*, vol. 25, no. 1, Art. no. 1, Jan. 2019, doi: 10.1038/s41591-018-0300-7.

[3] S. Abdulrahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, and M. Guizani, "A Survey on Federated Learning: The Journey From Centralized to Distributed On-Site Learning and Beyond," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5476–5497, Apr. 2021, doi: 10.1109/JIOT.2020.3030072.

[4] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," *ArXiv160205629 Cs*, Feb. 2017, Accessed: Mar. 10, 2022. [Online]. Available: http://arxiv.org/abs/1602.05629

[5] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.

[6] T. Yang *et al.*, "Applied Federated Learning: Improving Google Keyboard Query Suggestions," arXiv, arXiv:1812.02903, Dec. 2018. doi: 10.48550/arXiv.1812.02903.

[7] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang, "Federated Learning for Healthcare Informatics," *ArXiv191106270 Cs*, Aug. 2020, Accessed: Dec. 10, 2021. [Online]. Available: http://arxiv.org/abs/1911.06270

[8] I. Dayan *et al.*, "Federated learning for predicting clinical outcomes in patients with COVID-19," *Nat. Med.*, vol. 27, no. 10, pp. 1735–1743, Oct. 2021, doi: 10.1038/s41591-021-01506-3.

[9] M. J. Sheller *et al.*, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Sci. Rep.*, vol. 10, no. 1, p. 12598, Dec. 2020, doi: 10.1038/s41598-020-69250-1.

[10] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated Optimization in Heterogeneous Networks." arXiv, Apr. 21, 2020. Accessed: Jun. 05, 2022. [Online]. Available: http://arxiv.org/abs/1812.06127

[11] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2020," *CA. Cancer J. Clin.*, vol. 70, no. 1, pp. 7–30, 2020, doi: 10.3322/caac.21590.

[12] C.-C. Chang *et al.*, "Developing a Stacked Ensemble-Based Classification Scheme to Predict Second Primary Cancers in Head and Neck Cancer Survivors," *Int. J. Environ. Res. Public. Health*, vol. 18, no. 23, p. 12499, Nov. 2021, doi: 10.3390/ijerph182312499.

[13] C.-C. Chang and S.-H. Chen, "Developing a Novel Machine Learning-Based Classification Scheme for Predicting SPCs in Breast Cancer Survivors," *Front. Genet.*, vol. 10, 2019, Accessed: Jun. 06, 2022. [Online].

[14] W.-C. Ting, H.-R. Chang, C.-C. Chang, and C.-J. Lu, "Developing a Novel Machine Learning-Based Classification Scheme for Predicting SPCs in Colorectal Cancer Survivors," *Appl. Sci.*, vol. 10, no. 4, Art. no. 4, Jan. 2020, doi: 10.3390/app10041355.

[15] D. J. Beutel *et al.*, "Flower: A Friendly Federated Learning Research Framework," arXiv, arXiv:2007.14390, Mar. 2022. doi: 10.48550/arXiv.2007.14390.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," arXiv, arXiv:1708.02002, Feb. 2018. doi: 10.48550/arXiv.1708.02002.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, 2002, doi: 10.1613/jair.953.